UCSF
UC San Francisco Previously Published Works

Title

Does it Fit? Testing a Novel Pediatric Surgical Risk Calculator in Africa

Permalink

https://escholarship.org/uc/item/0rj6v787

Journal

Journal of Pediatric Surgery, 62(1)

ISSN

0022-3468

Authors

Zambeli-Ljepović, Alan
Negash, Samuel
Kinfe, Rahwa Amha
et al.

Publication Date

2026-10-01

DOI

10.1016/j.jpedsurg.2026.163524

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Does it Fit? Testing a Novel Pediatric Surgical Risk Calculator in Africa

Alan Zambeli-Ljepović [1,2] , Samuel Negash [2, 3], Rahwa Amha Kinfe [4], Samuel Mesfin Girma [4], 2 Architha Kannan [2], Chisom Udeigwe [5], Doruk Ozgediz [1,2] , Tihitena Negussie Mammo [5,6], Samuel Gashu [6], Felix Alakaloko [7] ,Justina Seyi-Olajide [7],Emmanuel Ameh [5]

[1] Department of Surgery, University of California San Francisco, USA; [2] Institute of Global Health Sciences, University of California San Francisco, USA; [3] Department of Surgery, Menelik II Hospital, Ethiopia; [4] Tikur Anbessa Specialized Hospital; [5] Department of Surgery, National Hospital, Nigeria; [6] Department of Surgery, Addis Ababa University, Ethiopia; [7] Department of Surgery, Lagos University Teaching Hospital, Nigeria

Abstract

Background: Low- and middle-income countries (LMICs), particularly Sub-Saharan Africa, experience a high burden of surgical disease and a disproportionately high postoperative morbidity and mortality. Although machine learning based pediatric surgical risk calculators (SRCs) have been developed to support clinical decision making, most are derived from high-income settings and may not generalize well to LMIC populations. We previously developed a pediatric surgical risk calculator using multinational LMIC data; however, its performance in independent African cohorts remains unknown.

Methods: We conducted an external validation study of a previously developed pediatric surgical risk calculator using prospective data from children undergoing surgery at four tertiary referral hospitals in Ethiopia and Nigeria between July 2024 and February 2025. Model performance was assessed by evaluating discrimination (area under the receiver operating characteristic curve) and calibration (slope and intercept). Predicted risk was stratified into low, intermediate, and high-risk categories to enhance clinical applicability. Model recalibration was performed to optimize performance in the external cohort.

Results: A total of 1,670 pediatric surgical patients were included from the four hospitals. In-hospital mortality was 6.8%. Mortality varied significantly by site, age, weight, sex, prematurity, preoperative sepsis, transfusion, supplemental oxygen use, antibiotic use, ASA class, type of surgery (emergent vs elective), need of ICU postoperatively and reoperation. The risk prediction model had good discrimination overall with an area under the curve (AUC) of 0.84, outperforming ASA classification alone (AUC = 0.77). Discrimination remained high for general pediatric surgery cases (AUC=0.83) and was highest for emergency cases (AUC = 0.89), but was low in neonates (AUC = 0.64). The model demonstrated systematic miscalibration, which improved following logistic calibration. Patients were stratified into meaningful risk categories, which reliably distinguishes between low-and-hgih risk patients (LR=0.1 and LR=5.4 respectively). However, 66.5% fall in the intermediate risk group which is not useful in decision making (LR=1.0).

Conclusion: The pediatric surgical risk calculator demonstrated good performance in the African context, with improved clinical utility following recalibration and risk stratification. Reduced performance in neonates suggests the need to incorporate additional risk factors and potentially separate models for this subgroup. Future efforts should focus on supporting the use of the SRC in risk communication, clinical decision making, and resource prioritization.

Keywords: machine learning algorithm; Mortality prediction; Pediatric surgery; global health; Global surgery

Introduction

There is a substantial unmet need for surgical care among children living in low-and-middle income countries (1). This burden is greatest in Sub-Saharan Africa where children comprise nearly half of the population (2). Children in these settings often present late with advanced disease, malnutrition and sepsis (3). In addition, there is a shortage of surgical care providers, along with limited perioperative and critical care infrastructure (4,5). These factors collectively contribute to disproportionately high rates of mortality, which can be up to seven times higher than those in high-income countries (6–8). Given this burden, efforts should focus on identifying and prioritizing those at highest risk of postoperative mortality in LMICs and mitigating this risk where possible (9).

Recent advances in machine learning offer a promising approach to this challenge by enabling the analysis of large-scale real-world data to predict health outcomes (10). Such approaches represent a significant advance over traditional methods, as they can identify highly complex patterns in data and generate accurate predictions without relying heavily on statistical assumptions (11). The application of these methods in surgical care has grown rapidly, with the development of numerous SRCs that support clinical decision-making by identifying high-risk patients who may benefit from enhanced monitoring and interventions (12–14). The approach has also been extended to pediatric populations (15–17). However, most of these tools have been trained using data from high income countries, raising concerns about their use in LMICs, which have different patient populations and healthcare infrastructure(18,19).

Our group developed a pediatric surgical risk calculator using data from a multinational perioperative database encompassing LMICs across Africa and Latin America (20). The calculator is based on a gradient boosting machine learning model and uses preoperative variables including patient demographics, comorbidities, procedural factors, and physiologic indicators to predict postoperative mortality. Initial tests demonstrated good predictive performance, however, prediction models may exhibit reduced performance when applied to populations outside their original development dataset (21). This is particularly relevant given our model was derived from pooled data across 14 countries with considerable heterogeneity (20). Therefore, external validation in an independent dataset is essential prior to widespread clinical adoption (21).

In this study, we externally validated this pediatric surgical risk calculator using prospective data from four hospitals in Ethiopia and Nigeria. We further performed recalibration to improve performance and examined risk stratification to enhance clinical utility. Our aim was to optimize the model's performance and determine its readiness for implementation in routine clinical practice.

Methods

Study Design and Setting

This study is an external validation of a recently developed Surgical Risk Calculator for children undergoing surgery in LMICs (20). We assessed discrimination, calibration, and clinical utility (using decision-curve analysis) to objectively evaluate the model's performance in a cohort of patients separate from the derivation cohort. This manuscript was prepared following the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) framework (22).

We collected data from four tertiary referral hospitals in Nigeria and Ethiopia, the two most populous countries in Africa, with populations of approximately 233 million and 132 million, respectively (23). In both countries, children represent a significant share of the population, with 41% in Nigeria and 39% in Ethiopia being under 15 years of age (24). In Ethiopia, data were collected from Tikur Anbessa Specialized Hospital (TASH) and Menelik II Comprehensive Specialized Hospital (MCSH). In Nigeria, participating centers were Lagos University Teaching Hospital (LUTH) and National Hospital Abuja (NHA). The hospitals were selected as the high-volume tertiary referral centers providing pediatric surgical care.

In Ethiopia, both hospitals have more than 800 beds and NICU services, however, dedicated pediatric ICU was only available at TASH. TASH has six pediatric surgeons and a longstanding pediatric surgery service performing approximately 80 procedures per month, including oncologic and complex congenital anomalies. It has established multidisciplinary teams as well as pediatric surgery residency and fellowship programs. MCSH has two pediatric surgeons and a more recently established service performing approximately 40 procedures per month, with trainees rotating from TASH.

In Nigeria, LUTH has more than 800 beds. Although the hospital has no standalone Pediatric ICU or NICU, the ICU has designated space for 2 children and neonates. It has a longstanding pediatric surgery service and training program with an operative volume of around 70 procedures per month . NHA has more than 400 beds, a NICU, and 2 operating rooms dedicated to pediatric surgery but does not have a Pediatric ICU. There are four pediatric surgeons and operative volume is approximately 40 per month.

Data Collection and Eligibility Criteria

We sequentially collected perioperative clinical information for every child (age <18 years) who underwent a surgical operation at the four aforementioned hospitals between July 2024 to February 2025. Given the relative rarity of the event of interest (inpatient mortality), we continued data collection until we observed at least 100 events, to ensure an adequate number of outcome events to evaluate the prediction model. We collected all variables used in SRC derivation, as described previously and shown in **Table 1**. We excluded patients with a missing discharge status (discharged home, left against medical advice, transferred, or deceased). Data were collected and stored using an encrypted electronic data capture system (REDCap) (25), hosted at the University of California, San Francisco (UCSF).

Mortality Risk Model and External Validation

The derivation and characteristics of the mortality risk prediction model have been described previously(20). Briefly, it is an extreme gradient boosting model, which, on cross-validation, had an area under the receiver-operator characteristic (ROC) curve (AUC) of 0.862 (95% CI 0.846 to 0.882) and calibration slope and intercept of 1.03 (95% CI 0.97 to 1.09) and 1.18 (95% CI 0.98 to 1.39), respectively.

To evaluate the SRC's performance in an external cohort, we assessed discrimination using ROC curve analysis and calibration using the slope and y-intercept of the curve representing observed versus predicted mortality rates by decile of predicted risk.

Following common practice in clinical risk prediction, we stratified predicted risk into three categories (low, intermediate, and high) – each with an associated interval likelihood ratio (LR) – to provide a more clinically useful output (26). Conceptually, we used low risk to indicate that it is safe to proceed with surgery and high risk to indicate that surgery, if performed at all, should not be performed without more significant perioperative support (further resuscitation, blood products, critical care resources). The intermediate risk category comprised patients for whom the SRC could not reliably predict risk. We set the lower risk threshold (between the low and intermediate risk categories) at the point where SRC sensitivity was 97~~9~~%, so that the likelihood of mortality in a patient classified as "low risk" would not exceed 3~~1~~%. We ~~set~~ iteratively tested varying ~~the~~ upper risk thresholds (between the intermediate and high risk categories) ~~at~~ until we found the value that maximized the clinical utility of resultant interval LRs (Supplement)~~the predicted risk which maximized the net benefit on decision curve analysis~~. We re-evaluated the AUC using the stratified risk model and ~~computed~~ reported interval LRs~~likelihood ratios (LRs)~~ for each risk category.

Because external validation revealed inadequate calibration, we logistically recalibrated the model to better align predicted probabilities with observed outcomes in our external cohort. We similarly stratified the recalibrated model into three risk categories and evaluated the same model performance metrics as for the original (unrecalibrated) model. All analyses were performed using R version 4.3.1 (27).

Ethical considerations

Ethical approval for this study was obtained from the institutional review committees at all four participating hospitals. The study was deemed exempt from institutional review at UCSF, as UCSF investigators did not have access to identifiable patient data. Documentation of this exemption is available upon request.

Results

Basic Demographic and Clinical Characteristics

A total of 1,670 pediatric surgical patients were included from the four hospitals : TASH (35%), LUTH (30%), MCSH (16%), and NHA (18%). The overall in-hospital mortality rate was 6.8% (n = 114). Mortality varied significantly by site: TASH (12.4%), LUTH (5.1%), NHA (4.3%), and

MCSH (0.7%) ($p < 0.001$) **(Table 1)**.

Children who died were significantly younger (median age newborn (IQR: 0.0–0.2) vs. 3.2 years (1.0–7.2), $p < 0.001$), had lower weight (3 kg (3–5) vs. 14 kg (9–22), $p < 0.001$), and were more likely to have been born premature (22% vs. 7.8%, $p < 0.001$). Female patients comprised 44% of deaths compared to 28% of survivors ($p < 0.001$).

Preoperative sepsis (11% vs. 2.6%, $p < 0.001$), transfusion (27% vs. 10%, $p < 0.001$), supplemental oxygen use (32% vs. 23%, $p = 0.037$), and antibiotic use (93% vs. 78%, $p < 0.001$) were more common among patients who died. Higher ASA class was also significantly associated with increased mortality ($p < 0.001$). General anesthesia was used in 93% of cases overall. There were no significant differences in anesthesia type or timing of surgery between groups.

The most common specific procedures were Laparotomy (8%), hypospadias correction (7%), and inguinal hernia repair (7%), with other common procedures comprising 4-6% of cases (**Figure 1)**. Mortality was higher following emergent surgery (34% vs. 15%, $p < 0.001$), and among patients transferred from another facility (89% vs. 78%, $p = 0.004$). Mortality was also higher among patients admitted to the ICU postoperatively (28% vs. 5.3%, $p < 0.001$) and among those who required but did not receive ICU care (70% mortality). Reoperation (8.1% vs 2.2%, $p = 0.035$) was also more frequent among those who died.

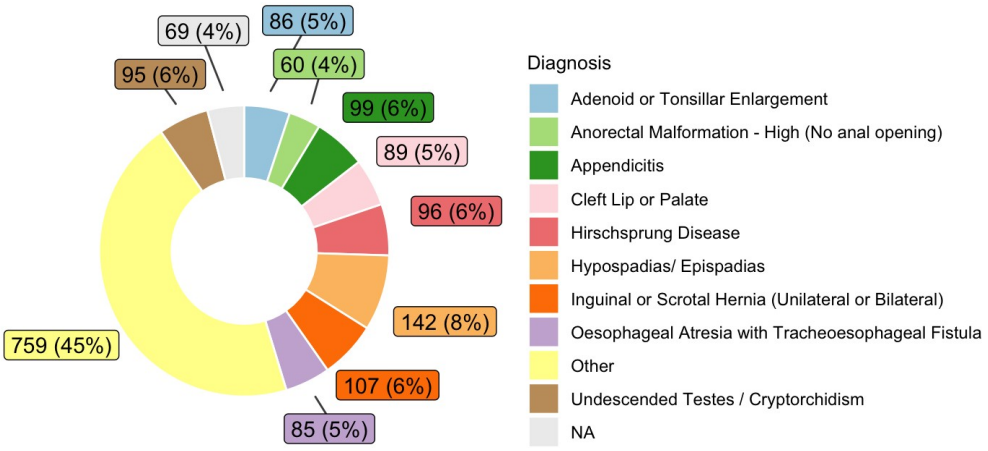
Table 1*. Baseline characteristics and clinical outcomes*

Characteristic	**All Patients**[1] N = 1,670	**Alive** N = 1,556[1]	**Deceased** N = 114[1]	**p-value**[2]
Hospital				<0.001
TASH	588 (35%)	515 (88%)	73 (12%)	
LUTH	509 (30%)	483 (95%)	26 (5%)	
NHA	299 (18%)	286 (96%)	13 (4%)	
MCSH	274 (16%)	272 (99%)	2 (1%)	
Demographics				
Age, years	3.0 (0.7, 7.0)	3.2 (1.0, 7.2)	0.0 (0.0, 0.2)	<0.001
Female Sex	490 (29%)	439 (90%)	51 (10%)	<0.001
Weight, kg	14 (8, 23)	14 (9, 24)	3 (3, 5)	<0.001
Prematurity	122 (8.9%)	98 (80%)	24 (20%)	<0.001
Emergent Operation	274 (16%)	234 (15%)	40 (35%)	<0.001
Time of Operation				0.168
Day	1,430 (87%)	1,333 (93%)	97 (7%)	
Night	134 (8.1%)	127 (95%)	7 (5%)	
Weekend	86 (5.2%)	76 (88%)	10 (12%)	
Transferred from another center	1,298 (79%)	1,196 (92%)	102 (8%)	0.003
Primary Team				<0.001
General Pediatric Surgery	1,313 (80%)	1,206 (92%)	107 (8%)	
Other	328 (20%)	324 (99%)	4 (1%)	
Anesthesia type				
Local Anesthesia	24 (1.4%)	21 (88%)	3 (13%)	0.222
Regional Anesthesia	99 (5.9%)	93 (94%)	6 (6%)	>0.999
General Anesthesia	1,558 (93%)	1,448 (93%)	110 (7%)	0.240
MAC Anesthesia	46 (2.8%)	45 (98%)	1 (2%)	0.366

No Anesthesia	11 (0.7%)	11 (100%)	0 (0%)	>0.999
ASA Classification				<0.001
ASA I	752 (46%)	749 (99%)	3 (1%)	
ASA II	779 (48%)	698 (90%)	81 (10%)	
ASA III	88 (5.4%)	67 (76%)	21 (24%)	
ASA IV	17 (1.0%)	9 (53%)	8 (47%)	
ASAV	2 (0.1%)	1 (50%)	1 (50%)	
Other Pre-operative factors				
Temperature	36.5 (36.3, 36.8)	36.5 (36.3, 36.8)	36.5 (36.4, 36.8)	0.170
Sepsis	52 (3.1%)	40 (77%)	12 (23%)	<0.001
Antibiotics	1,305 (79%)	1,199 (92%)	106 (8%)	<0.001
Transfusion	192 (12%)	161 (84%)	31 (16%)	<0.001
Supplemental Oxygen	408 (25%)	371 (91%)	37 (9%)	0.042
Safety Checklist	1,567 (94%)	1,458 (93%)	109 (7%)	0.677
Post-operative factors				
Temperature	36.7 (36.3, 36.9)	36.7 (36.3, 36.9)	36.7 (36.3, 36.9)	0.584
ICU admission	112 (6.8%)	80 (71%)	32 (29%)	<0.001
ICU Needed but Unavailable	142 (9.2%)	85 (60%)	57 (40%)	<0.001
Antibiotics	1,205 (73%)	1,099 (91%)	106 (9%)	<0.001
Sepsis	135 (8.1%)	70 (52%)	65 (48%)	<0.001
Wound Infection	84 (5.1%)	74 (88%)	10 (12%)	0.069
Reoperation	43 (2.6%)	34 (79%)	9 (21%)	0.002

[1]n (%); Median (Q1, Q3)

[2]Fisher's exact test for categorical variables; Wilcoxon rank sum test for continuous variables



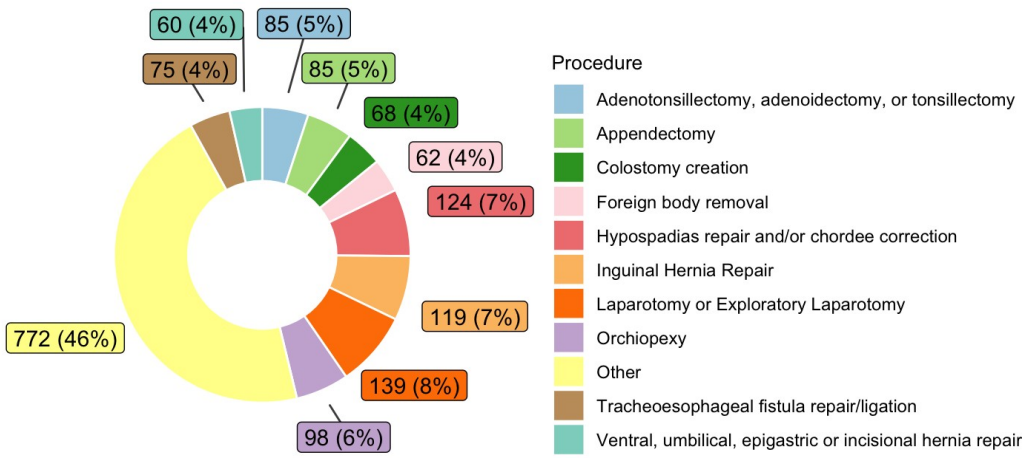


***Figure 1**. Top 10 most common diagnoses (A) and surgical procedures (B).*

Model Performance: Discrimination, Calibration, and Clinical Utility

The model had an area under the ROC curve (AUC) of 0.842 (95% CI: 0.811–0.873) in the full sample **(Figure 2**). Performance exceeded that of ASA classification alone (AUC = 0.766, p = 0.0005) . Discrimination remained high in emergent cases (AUC = 0.885) and general pediatric surgery (AUC = 0.825), but was lower in neonates (AUC = 0.637).

The model showed systematic miscalibration with a calibration slope of 0.654 (95% CI: 0.545–0.765) and calibration-in-the-large of 1.923 (95% CI: 1.685–2.151) (**Figure 3**)

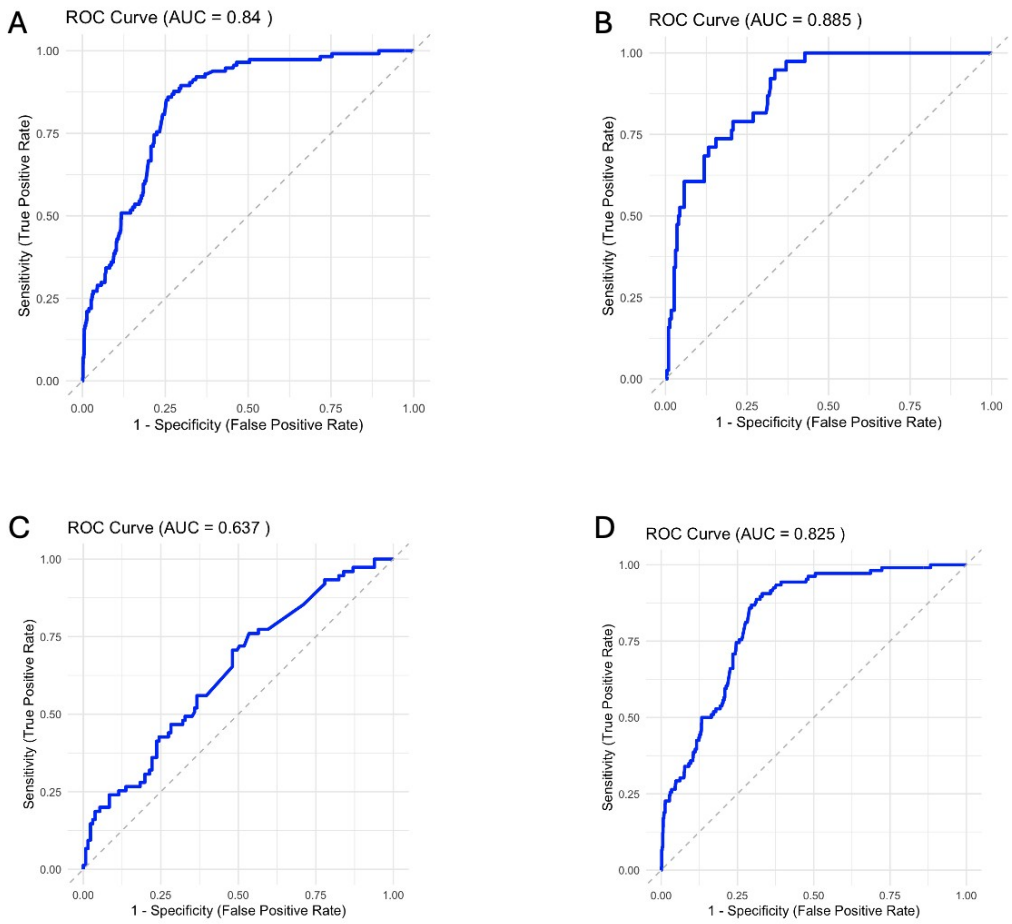


***Figure 2**. Receiver-operator characteristic curve of all patients (A), emergent cases (B), neonates (C) and general pediatric surgery (D)*

To assess clinical utility, we plotted a decision curve for the predictive model (**Supplement**), which demonstrated a net clinical benefit at risk thresholds of ≥5%. Using this risk level as the upper threshold and the risk level corresponding to 99% model sensitivity as the lower threshold, we stratified the cohort into three risk groups: low (predicted probability <0.01): LR^- = 0.01–0.59, intermediate,(0.01–0.05): LR = 0.84 (95% CI: 0.74–0.94) and high (>0.05): LR^+ = 5.24 (95% CI: 3.68–7.46). Notably, this placed 1386 (83.0%) of patients in the intermediate risk category. (**Figure 3)**

Recalibration

Logistic recalibration was used to update the model intercept and slope using logistic regression. We found improved model performance after recalibration **(Table 2)**, achieving a calibration slope of 1.000 and calibration-in-the-large of 0, while maintaining discrimination (AUC unchanged at 0.842). More importantly, recalibration reduced the proportion of patients in

the intermediate risk category to 1112 patients (66.5%) while preserving clinically meaningful interval likelihood ratios (**Figure 4**)

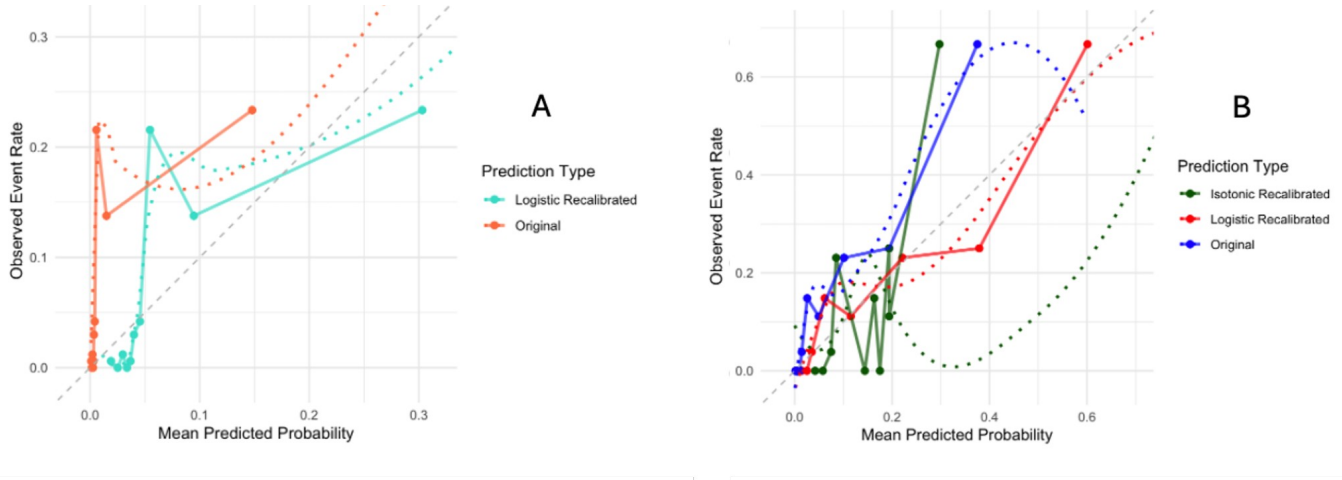


Figure 3 *(Re)Calibration plot for all patients (A), and emergent cases (B)*

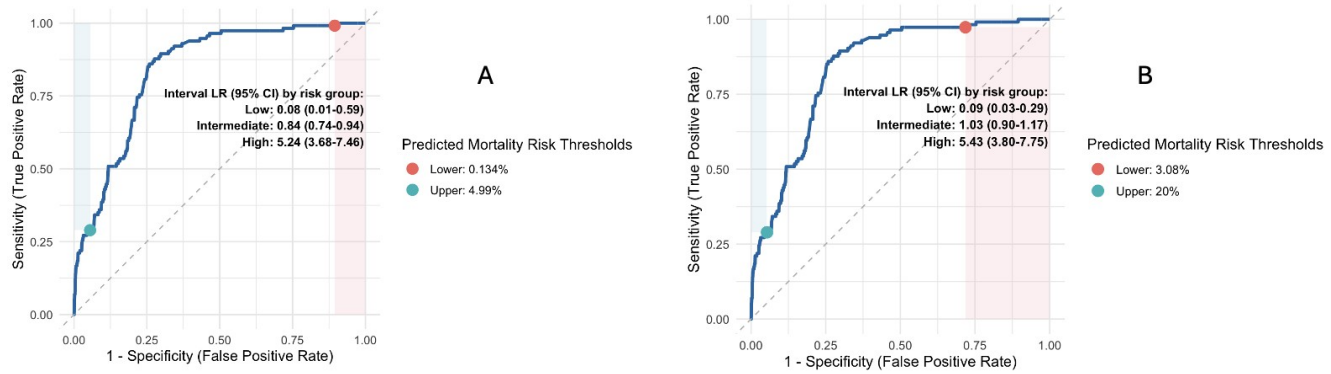


Figure 4*. Risk stratification and associated interval likelihood ratios for (A) model prior to recalibration, using conservative thresholds, (B) model after logistic recalibration, using thresholds that maximize utility of iLR*

***Table 2**. Model Re (calibration)*

	All cases		Emergent cases		
Parameter	**Original**	**Recalibrated (Logistic)**	**Original**	**Recalibrated (Logistic)**	**Recalibrated (Isotonic)**
ROC [95% CI]	0.840 [0.809, 0.871]	0.840 [0.809, 0.871]	0.885 [0.838, 0.933]	0.885 [0.838, 0.933]	0.641 [0.545, 0.737]
Calibration Slope [95% CI]	0.637 [0.530, 0.745]	1.000 [0.833, 1.170]	1.012 [0.726, 1.353]	1.000 [0.717, 1.336]	0.077 [0.021, 0.138]
Intercept [95% CI]	0.430 [-0.065, 0.921]	0.000 [-0.430, 0.425]	0.959 [0.235, 1.724]	-0.000 [-0.519, 0.529]	-1.304 [-1.784, -0.862]
Calibration-in-the-Large [95% CI]	1.855 [1.617, 2.084]	0.000 [-0.209, 0.199]	0.934 [0.513, 1.335]	-0.000 [-0.422, 0.402]	-2566202051822478.500 [NA, NA]
Mean Error (Bias)	-0.05	0	-0.067	0	0
Mean Absolute Error	0.075	0.112	0.153	0.172	0.212
RMSE	0.245	0.238	0.311	0.294	0.366
Brier Score (p-value)	0.060 (p=0.000)	0.057 (p=0.000)	0.097 (p=0.000)	0.086 (p=0.000)	0.134 (p=0.000)

Discussion

We externally validated a pre-existing model for predicting mortality following children's surgery in LMICs (20) using prospective data from Nigeria and Ethiopia. The model demonstrated good discrimination overall (AUC = 0.84), with especially strong discrimination in emergency cases (AUC = 0.89), outperforming ASA classification alone (AUC = 0.77). Calibration analysis revealed systematic patterns that were addressed through logistic recalibration. We further enhanced clinical utility through three level risk stratification.

Our model's discrimination on external validation (AUC 0.84) positions it within the range of clinically acceptable performance. The model had excellent performance during original testing, with internal validation (AUC 0.95), comparable to SRCs developed in high income countries, including NZRISK-pediatric from New Zealand (AUC = 0.95) and ACS NSQIP from the United States (AUC = 0.96) (16,17, 20). However, the performances in these models were not externally validated. In addition these models were developed using national databases within relatively uniform healthcare systems while our model incorporated data from multiple countries with diverse healthcare systems. The lower discrimination observed during external validation likely reflects differences in case mix and contribution from unmeasured factors such as delays and resource constraints (9).

Systematic miscalibration is a common challenge when applying predictive models across different health care contexts. We found the classic pattern of overestimating risk at low probabilities while underestimating risk at higher probability levels, consistent with prior validation studies (9). However, there was also underestimation of risk for some patients at low predicted probabilities, which is particularly important to consider in clinical application. Logistic recalibration improved these discrepancies, while preserving discrimination. This finding indicates that, prior to implementation, clinical predictive models should be tested in and recalibrated to the setting in which they're being applied (29).

To optimize clinical utility, we created a three-level risk stratification (low, intermediate, high) of model outputs. By consensus, we prioritized a highly sensitive threshold between the low- and intermediate-risk groups that minimizes the risk of underestimating mortality. In setting the threshold between the intermediate- and high-risk groups, we prioritized net benefit. This conservative approach reliably distinguishes between low- and high-risk patients (LR = 0.1 and LR=5.4, respectively) but comes at the expense of classifying two-thirds of patients as intermediate risk, where the LR ~1 is not helpful in decision making. We stress that, regardless of risk stratification provided by the model, clinical judgment remains paramount in determining risk and mitigating risk through appropriate interventions.

Importantly, the model's performance varied across patient subgroups. Discrimination was highest in emergent cases (AUC = 0.89) and general pediatric surgery (AUC = 0.83), but was substantially limited in neonates (AUC = 0.64). The strong discrimination in emergent cases is particularly relevant given the need for time sensitive decisions. However, reduced performance in neonates represents an important limitation, and risk estimates in this subgroup should be interpreted with caution. The lower discrimination among neonates may reflect their distinct

physiology and disease presentations, their relative underrepresentation in the development and validation cohorts, or neonatal specific risk factors that may not be fully captured by the current SRC (30). Future studies should consider developing models using dedicated neonatal datasets and incorporating more granular predictors to improve risk prediction in this population.

The most significant limitation was insufficient sample size for robust site specific recalibration. While our overall cohort of 1,670 patients with 114 events provided adequate power for external validation, meaningful site specific analysis would require larger sample sizes at each institution. This limits our ability to account for institutional variations in cases, treatments and outcomes that may influence model performance.

Looking ahead, the SRC could support clinical decision-making and resource allocation decisions in settings where resources are limited. Risk estimates may help identify high risk children who may benefit from prioritization of scarce resources such as blood transfusions and ICU beds or escalation of care by transfer to a higher facility or operative intervention. Clinicians can also use the tool to facilitate informed consent conversations with caregivers, inform quality improvement efforts, and at the population level to better characterize the risk profiles of children undergoing surgery. Future implementation studies should identify the optimal way to integrate this tool into clinical care and assess its utility in clinical decision-making.

Conclusion

This study presents the first external validation of a pediatric surgical risk calculator in Africa. The SRC demonstrated good predictive performance across tertiary hospitals in Ethiopia and Nigeria, with improved performance after recalibration. Risk-stratification meaningfully classified one-third of the children to low and high risk groups. Subgroup analyses showed that performance remained high among emergent and general pediatric surgery cases but was substantially lower among neonates, highlighting an important limitation of the current model.

Acknowledgments

We acknowledge the data collectors (Beimnet Girma Negussie, Babatunde Ademola Dickson, Olapeju Ishola, and Nancy Nneka Ukwu), without whose diligence the development of the risk calculator and this validation would not have been possible. We also acknowledge Michael Kohn at the University of California, San Francisco, for his advice on the methods for evaluating risk prediction models, and KidsOR, whose philanthropy has increased capacity for children's surgery worldwide and facilitated an unprecedented quality of pediatric surgical research.

References

1. Mullapudi B, Grabski D, Ameh E, Ozgediz D, Thangarajah H, Kling K, et al. Estimates of number of children and adolescents without access to surgical care. Bull World Health Organ. 2019 Apr 1;97(4):254–8. doi:10.2471/BLT.18.216028 PubMed PMID: 30940982; PubMed Central PMCID: PMC6438256.

2. Bandyopadhyay S, Philipo GS, Banu T, Bokhary Z, Lakhoo K. Global Initiative for Children's Surgery (GICS): a decade in review. World J Pediatr Surg. 2025 Oct 2;8(4):e001084. doi:10.1136/wjps-2025-001084 PubMed PMID: 41050908; PubMed Central PMCID: PMC12496102.

3. Landrum KR, Pence BW, Shrime MG, Rice HE, Wright NJ, Habermann A, et al. Delayed primary surgery and outcomes in children with gastrointestinal anomalies in 264 hospitals and 74 countries. Am J Epidemiol. 2025 Nov 4;194(11):3332–47. doi:10.1093/aje/kwaf160

4. Derbew M. Pediatric surgery in Eastern Africa: The unmet need. J Pediatr Surg. 2019 Jan;54(1):21–6. doi:10.1016/j.jpedsurg.2018.10.028 PubMed PMID: 30454942.

5. Hsu PJ, Carroll M, Zambeli-Ljepovic A, Olatunji-Banire BT, Mathew P, Axt J, et al. Current Landscape of Children's Surgery in Africa: A Multicenter Analysis of 16,000 Cases. World J Surg. 2025 Nov;49(11):3163–72. doi:10.1002/wjs.70066 PubMed PMID: 40913246; PubMed Central PMCID: PMC12582145.

6. Wright NJ, Leather AJM, Ade-Ajayi N, Sevdalis N, Davies J, Poenaru D, et al. Mortality from gastrointestinal congenital anomalies at 264 hospitals in 74 low-income, middle-income, and high-income countries: a multicentre, international, prospective cohort study. The Lancet. 2021 Jul 24;398(10297):325–39. doi:10.1016/S0140-6736(21)00767-4 PubMed PMID: 34270932.

7. Paediatric surgical outcomes in sub-Saharan Africa: a multicentre, international, prospective cohort study. BMJ Glob Health. 2021 Sep 2;6(9). doi:10.1136/bmjgh-2020-004406 PubMed PMID: 10.1136/bmjgh-2020-004406.

8. Torborg A, Meyer H, Fiky ME, Fawzy M, Elhadi M, Ademuyiwa AO, et al. Outcomes after surgery for children in Africa (ASOS-Paeds): a 14-day prospective observational cohort study. The Lancet. 2024 Apr 13;403(10435):1482–92. doi:10.1016/S0140-6736(24)00103-X PubMed PMID: 38527482.

9. Kluyts HL, le Manach Y, Munlemvo DM, Madzimbamuto F, Basenero A, Coulibaly Y, et al. The ASOS Surgical Risk Calculator: development and validation of a tool for identifying African surgical patients at risk of severe postoperative complications. Br J Anaesth. 2018 Dec;121(6):1357–63. doi:10.1016/j.bja.2018.08.005 PubMed PMID: 30442264.

10. Machine Learning in Medicine | New England Journal of Medicine [Internet]. [cited 2026 May 10]. Available from: https://www.nejm.org/doi/full/10.1056/NEJMra1814259

11. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial Intelligence and Surgical Decision-making. JAMA Surg. 2020 Feb 1;155(2):148–58. doi:10.1001/jamasurg.2019.4917 PubMed PMID: 31825465; PubMed Central PMCID: PMC7286802.

12. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, Motaei A, Madkour M, Pardalos PM, et al. MySurgeryRisk: Development and Validation of a Machine-Learning Risk Algorithm for Major Complications and Death after Surgery. Ann Surg. 2019 Apr;269(4):652–62. doi:10.1097/SLA.0000000000002706 PubMed PMID: 29489489; PubMed Central PMCID: PMC6110979.

13. Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR. Utilizing Machine Learning Methods for Preoperative Prediction of Postsurgical Mortality and Intensive Care Unit Admission. Ann Surg. 2020 Dec;272(6):1133. doi:10.1097/SLA.0000000000003297

14. Kilic A, Goyal A, Miller JK, Gjekmarkaj E, Tam WL, Gleason TG, et al. Predictive Utility of a Machine Learning Algorithm in Estimating Mortality Risk in Cardiac Surgery. Ann Thorac Surg. 2020 Jun 1;109(6):1811–9. doi:10.1016/j.athoracsur.2019.09.049

15. Cooper JN, Minneci PC, Deans KJ. Postoperative neonatal mortality prediction using superlearning. J Surg Res. 2018 Jan 1;221:311–9. doi:10.1016/j.jss.2017.09.002

16. Bertsimas D, Li M, Zhang N, Estrada C, Scott Wang HH. High-performance pediatric surgical risk calculator: A novel algorithm based on machine learning and pediatric NSQIP data. Am J Surg. 2023 Jul;226(1):115–21. doi:10.1016/j.amjsurg.2023.03.009 PubMed PMID: 36948897.

17. Mistry M, Boyle L, Taylor E, Campbell D. Development and validation of a national perioperative mortality risk model for pediatric surgery: The New Zealand pediatric surgical risk tool (NZRISK-pediatric): NZRISK-Pediatric. J Pediatr Surg. 2023 Mar;58(3):524–31. doi:10.1016/j.jpedsurg.2022.07.017 PubMed PMID: 35970677.

18. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithm fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023 Jun;7(6):719–42. doi:10.1038/s41551-023-01056-8 PubMed PMID: 37380750; PubMed Central PMCID: PMC10632090.

19. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447–53. doi:10.1126/science.aax2342 PubMed PMID: 31649194.

20. Eyler Dang L, Klazura G, Yap A, Ozgediz D, Bryce E, Cheung M, et al. Development and Internal-External Validation of a Post-Operative Mortality Risk Calculator for Pediatric Surgical Patients in Low- and Middle- Income Countries Using Machine Learning. J Pediatr Surg. 2024 Dec;59(12):161883. doi:10.1016/j.jpedsurg.2024.161883 PubMed PMID: 39317568.

21. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016 Jun 22;353:i3140. doi:10.1136/bmj.i3140 PubMed PMID: 27334381.

22. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015 Jan 6;13(1):1. doi:10.1186/s12916-014-0241-z

23. World Bank Open Data [Internet]. [cited 2026 May 10]. World Bank Open Data. Available from: https://data.worldbank.org

24. World Bank Open Data [Internet]. [cited 2026 May 10]. World Bank Open Data. Available from: https://data.worldbank.org

25. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019 Jul;95:103208. doi:10.1016/j.jbi.2019.103208 PubMed PMID: 31078660; PubMed Central PMCID: PMC7254481.

26. Kohn MA, Newman TB. Visualizing the value of diagnostic tests and prediction models, part III. Numerical example with discrete risk groups and miscalibration. J Clin Epidemiol. 2025 May;181:111691. doi:10.1016/j.jclinepi.2025.111691

27. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

28. Kraemer K, Cohen ME, Liu Y, Barnhart DC, Rangel SJ, Saito JM, et al. Development and Evaluation of the American College of Surgeons NSQIP Pediatric Surgical Risk Calculator. J Am Coll Surg. 2016 Nov;223(5):685–93. doi:10.1016/j.jamcollsurg.2016.08.542 PubMed PMID: 27666656.

29. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019 Dec 16;17(1):230. doi:10.1186/s12916-019-1466-7

30. Cooper JN, Minneci PC, Deans KJ. Postoperative neonatal mortality prediction using superlearning. J Surg Res. 2018 Jan 1;221:311–9. doi:10.1016/j.jss.2017.09.002 PubMed PMID: 29229144.

31. Raymond BL, Wanderer JP, Hawkins AT, Geiger TM, Ehrenfeld JM, Stokes JW, et al. Use of the American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator During Preoperative Risk Discussion: The Patient Perspective. Anesth Analg. 2019 Apr;128(4):643–50. doi:10.1213/ANE.0000000000003718 PubMed PMID: 30169413.